DOCUMENT RESUME

ED 023 564 RE 001 475

By-Herber, Harold L.

Teaching Reading through Seventh Grade Science Content.

Sewanhaka Central High School District Number 2, Nassau County, N.Y.

Note-30p.

EDRS Price MF-$0.25 HC-$1.60

Descriptors-*Content Reading, *Developmental Reading, *Reading Achievement, Reading Material Selection, *Reading Programs, *Reading Research, Reading Skills, Supplementary Reading Materials

The effectiveness of teaching specific, related reading skills within a content subject was tested in seventh-grade science classes. Comparisons were made among seventh graders who were taught no reading skills in conjunction with a content subject, seventh graders who received instruction in reading skills within the social studies content, and seventh graders who received instruction in reading skills within the science content. The achievement of the three groups was compared according to performance on the Iowa Silent Reading Test, the New York State Science Survey, and the New York State Social Studies Survey. Samples were drawn from the total population of approximately 2,000 students and evaluated for reliability. The level of significant achievement gain between pretest and post-test scores was determined, and the amount of growth from group to group was evaluated. Results indicated that the reading, science, and social studies instructional programs produced significant gains regardless of the developmental reading program, but growth in subject content achievement was measureably affected by this reading instruction. Teacher experience in reading skills instruction may have been a factor. An appendix gives detailed, independent corroboration of the data processing. (WB)

A-51-61

ED023564
RE 001 475

SEWANHAKA CENTRAL HIGH SCHOOL DISTRICT

A Report of the Experimental Program

for the school year 1961-62

TEACHING READING THROUGH SEVENTH GRADE SCIENCE CONTENT

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Dr. Harold L. Herber, Principal Investigator

Statement of the Problem Under Study

The purpose of this study is to investigate the validity and efficacy of a district wide developmental reading program in which science reading skills are taught through the content of seventh grade science, by regular classroom teachers, with no instructional time added to the course, and with no time set aside for reading instruction per se within the course.

In recent years there has been an increasing awareness of the necessity for providing instruction in reading to students in secondary schools. In addition, it has been urged by reading authorities that regular classroom teachers be responsible for teaching reading skills pertinent to their subject and grade level. Considerable research has been conducted to determine the effect of teaching reading and study skills in content areas. It has been demonstrated that such instruction improves understanding of the subject content and improves achievement in reading subject matter materials.

There is little evidence that the procedures advocated in research findings have been made practical by applying them to entire schools or districts. Such a program requires that materials be designed and teaching techniques be developed which would serve the individual ability and achievement levels of all participating students. This study is designed to make such a practical application of research findings.

Description of the Year's Work

- The Program of Instruction -

The actual beginning of the study occurred during the spring of 1961 when four teachers were appointed to serve in a summer workshop. These teachers were to write the materials to be used in the program. In discussions preparatory to the summer's work, the basic foundations of the program were established according to which the materials would be developed. It was agreed that when students are adequately prepared for the reading of any materials required in a subject, they will read with efficiency and understanding. It was established that this type of preparation should include adequate motivation, proper vocabulary study, a review of previously studied material to provide a context for new concepts, a development of anticipation of what might and what should be learned next in the course of a unit, and the giving of directions as to how and why the material should be read.

It was felt that subsequent to this preparation, as he actually read the material, the student could practice his reading skills guided by the special materials that were to be prepared.

It was decided that the materials which were to be written should follow the curriculum guide for seventh grade science in Central High School District No. 2. The materials were to be designed to provide for the development of science vocabulary and appropriate reading and study skills. Further, they were to be designed to meet varied ability and achievement levels. It was also decided that

these materials would be available in quantity to all teachers through the district's Instructional Materials Center.

During the summer of 1961, three seventh grade science teachers under the chairmanship of a reading specialist wrote the materials for the program. The committee examined each unit contained in the curriculum guide and identified the basic vocabulary related to each. These words became the source for a series of vocabulary exercises designed to develop familiarity with the words and their concepts. These vocabulary materials could be used by any teacher regardless of the textbook being used in the class since the words were basic to the curriculum and were common to all related textbooks.

The problem of providing for practice of appropriate reading and study skills on subject content material was a difficult one. Five different textbooks were being used within the district, making it impossible to use textbooks as the basis for practicing on the reading and study skills. Therefore, the committee identified sets of supplementary materials which were appropriate to the various units recommended in the curriculum guide. Worksheets were designed to be used by students so that as they read the supplementary materials they would be guided in practicing on the particular reading and study skills essential to the securing of the meaning of the text. The teacher then would be expected to teach so that students' ability to apply specific skills while reading the supplementary material would be transferred to their reading of the textbook being used in the class or to any other reading material related to the subject. It seemed reasonable to believe that this procedure would provide for enrichment of the course as well as the development of subject-related reading and study skills.

Even when quality materials have been written, however, the problem remains as to how to make them readily available to teachers and the students in each classroom. It is unrealistic to expect classroom teachers to cut stencils and provide materials for their students in the quantity essential in such a program. Therefore, it was determined that through the Instructional Materials Center, all materials written by the committee during the summer of 1961 would be available in quantity to each seventh grade science teacher. Each teacher had a complete set of materials which served as an order catalogue. There were more materials available in each unit than any one teacher could or would wish to use. The teachers were urged to be selective, basing choices of material on their estimate of the relative value of the particular unit in question, the skills and subject-content needs of their students, and upon the ability and achievement levels of their students. These materialswere not to displace other material that had been successful in developing proper understanding of subject matter in the past years. Old and new materials were to be combined in a concerted effort to develop both skills and understandings.

An orientation program was necessary to acquaint the teachers with the various teaching techniques that could be applied to these materials within this program and with the purposes and procedures relative to the program. This orientation took the form of an in-service course. It was conducted by the reading specialist who had acted as chairman of the group that wrote the materials. The teachers were urged to be aware of the fact that theywere not becoming "reading

teachers". They were seventh grade science teachers and this program was designed to help them teach seventh grade science in such a manner that the achievement needs of the students might be met. It was pointed out that the program was designed in such a way that skills and understandings could be developed simultaneously.

To give further assistance to the teachers involved in the program, the reading specialists in each building were to serve as consultants. They gave assistance through demonstrations, observations, consultations, and preparation of additional material as needed.

- The Research Design -

During the school year of 1960-61, a program similar to that which has been described above for seventh grade science was operative in seventh grade social studies. During the 1959-60 school year, there was no such program operative in the district. It was felt that the existence of the seventh grade social studies program provided a unique opportunity to measure not only the effectiveness of developing reading and study skills through science content, but it also provided opportunity to compare the results of a program operative in two subject areas with a program operative in one subject, with no program, all at the same grade level.

Therefore, it was determined that achievement of students in the seventh grade in the 1959-60 school year would provide the basic reference point. To this would be compared the achievement of students in the seventh grade during the 1960-61 school year when the program was operative in social studies. Also compared would be the achievement of students in the seventh grade in the 1961-62 school year when the program was operative in science as well as in social studies. The achievement of these three groups would be compared according to their performance on three measures: The Iowa Silent Reading Test; The New York StateScience Survey; and The New York StateSocial Studies Survey.

Miss Paula Miller, District Research Psychologist, and Dr. Howard Graves, Administrative Intern for Curriculum and Instruction, established the research design - in consultation with the writer and with STACO statistical consultants - and processed the data according to the design. Subsequently the data processing was corroborated independently by STACO. STACO's report appears in the appendix.

Since there were approximately 2000 students in the total population of each of the three groups, it was decided to select a representative sample from each group in order to facilitate the processing of the data. Accordingly, for each group, an independent proportional sample was selected, with replacement, from the five junior high schools in the district. The design called for an evaluation of the reliability of the samples for each achievement test and for the California Test of Mental Maturity. The standard score for the total test score on the Iowa was collected for each student in each group. The raw score for the total test score on both the Science Survey and the Social Studies Survey was collected for each student in each group.

The design called for the finding of the level of significance of the gain in achievement by each group as indicated by the change from pre-test score to post-test score on each of the above mentioned achievement tests. It also called for a determination of the significance of the amount of growth of these gains from group to group, using the 1959-60 group as the base line.

The set of regressions of pre-test on IQ and post-test on IQ for the three groups for all achievement tests was used in the analysis. Comparisons were made by the slopes and the intercepts. A detailed explanation of the statistical processes involved may be found in the appendix.

The 95% level of significance was established as the accepted level. Levels below the 95% were not accepted; however, depending upon proximity to the 95% point, some levels were considered indicative.

Analysis of the data

- Findings -

Pre to post-test achievement gain on the Iowa Silent Reading Test was significant for each of the three groups: 1959-60, 99.98% level amounting to 5.398 standard score points; 1960-61, 99.99% level amounting to 7.88 standard score points; 1961-62, 99.99+% level amounting to 10.41 standard score points. See Table I.

Pre to post-test achievement gain on the New York State Social Studies Survey was significant for each of the two groups for which complete data was available: 1959-60, 99.98% level amounting to 12.83 raw score points; 1961-62, 99.99+% level amounting to 15.97 raw score points. See Table IV.

Pre to post-test achievement gain on the New York State Science Survey was significant for each of the two groups for which complete data was available: 1959-60, 99.98% level amounting to 10.16 raw score points; 1961-62, 99.99 % level amounting to 12.03 raw score points. See Table V.

Comparing the relationship of achievement to ability, no differences were found for any of the three groups on any of the three measures on either pre or post-test data.

Achievement of the three groups on the Iowa Silent Reading Test was compared to determine the significance of the growth in the amount of pre to post-test gain experienced from group to group:

> Growth in the gain of the 1961-62 group compared to the gain of the 1959-60 group was significant at the 99% level and amounted to 5.01 standard score points. See Tables II and III.

Growth in the gain of the 1960-61 group compared to the 1959-60 group was not significant (77%) since it did not reach the accepted 95% level. The growth amounted to 2.48 standard score points. See Tables II and III.

Growth in the gain of the 1961-62 group compared to the 1960-61 group was not significant (80%) since it did not reach the accepted 95% level. The growth amounted to 2.53 standard score points. See Tables II and III.

Achievement of the two groups for which New York State Science Survey Test scores were available was compared to determine the significance of the growth in the amount of pre to post-test gain experienced from group to group:

Growth in the gain of the 1961-62 group compared to the 1959-60 group was not significant (40%) since it did not reach the accepted 95% level. The growth amounted to 1.87 raw score points. See Tables VI and VII.

Achievement of the two groups for which New York State Social Studies Survey Test scores were available was compared to determine the significance of the growth in the amount of pre to post-test gain experienced from group to group:

Growth in the gain of the 1961-62 group compared to the 1959-60 group was not significant (66%) since it did not reach the accepted 95% level. The growth amounted to 3.14 raw score points. See Tables VI and VII.

- Conclusions -

Whether the developmental reading program was operative in no subject area (1959-60), or was operative in social studies (1960-61), or was operative in both social studies and science (1961-62), the ability-achievement relationship remained constant for all measures.

Whether the developmental reading program was operative in no subject area, or was operative in social studies or was operative in both social studies and science, the pre to post-test achievement gain was significant on all three measures.

One finds no significant growth on the Iowa Silent Reading Test by comparing gains in achievement of the group when no developmental reading program was operative with gains of the group when the program was operative in social studies. However, this growth is indicative (77%). Lack of scores on the two survey tests for 1960-61 prevents similar comparisons of growth based on those two measures.

Comparing gains in achievement of the group when the program was operative in social studies with the group when the program was operative in social studies and science, one finds no significant growth on the Iowa Silent Reading Test. Again, however, this growth is indicative (80%). Lack of scores on the two survey tests for 1960-61 prevents similar comparisons of growth based on those two measures.

Comparing gains in achievement of the group when no developmental reading program was operative with gains of the group when the program was operative in both social studies and science, one does find significant growth on the Iowa Silent Reading Test (99%).

One finds no significant growth (40%) when comparing gains in achievement on the New York State Science Survey for these two groups.

Similarly, one finds no significant growth (66%) for these two groups on the New York State Social Studies Survey.

- Interpretation -

It is evident from the data that instruction in science, social studies, and reading was of sufficient quality to produce significant gains in achievement before the existence of the developmental reading program. The logical question, then, is: "Why have a developmental reading program?" "Of what value is it to have such a program?"

To answer the question, one must assume that when no reading program was operative, regular instructional procedures and experiences were responsible for the significant gain in achievement. It also must be assumed that the institution of any variance in teaching methods or materials in these two areas would have a measureable effect on student achievement. Thirdly, it must be assumed that the district has not reached the epitome of student achievement. Therefore, one can justifiably expect that the students at a given grade level and in a given subject area should achieve significantly better than the students of previous years at the same level. This, of course, would require that the IQ be factored out so variance in ability of groups would not effect an estimate of the significance of such growth.

An examination of the growth in achievement gains from year to year is important, then, so as to note improvement in the quality of education. Tables II and VI indicate the degree to which this improved quality is being realized. Since one may assume that the instructional variables changed in the social studies and science courses in question were due to the developmental reading program, one may assume that improvement or regression in quality of education is due to the methods and materials which are incorporated in the program and applied by the individual teachers in the program.

The reason for having such a program, then, is to improve constantly the quality of education, even though one could be satisfied that the status quo is producing significant pre to post-test achievement gains.

The only instance where significant growth in pre to post-test achievement gains occurred, was over a two-year period of time, in reading achievement, as measured by the Iowa Silent Reading Test. This two-year span covered the year in which the program was operative in any subject area (1959-60) and the year in which the program was operative in both social studies and science (1961-62). In no other between-group comparisons were there significant growths in achievement gains on any of the three measures.

One can conclude, therefore, that having a program operative in social studies and science does produce gains in reading achievement significantly greater than when no such program is operative in either area.

Because of lack of test data for the 1960-61 group, it is possible to measure growth in subject-content achievement gain only between the group which did not participate in a developmental reading program (1959-60) and the group which participated in a program in both social studies and science (1961-62). Table VI indicates the fact that there was no significant growth between these two groups on either the social studies or the science survey tests. This means that in terms of attaining significant growth in subject-content achievement gain, the existence of the developmental reading program in two subject areas made no difference.

A careful examination of Tables II and III reveals some very interesting trends, however. Though the significance in the growth of the gains in reading achievement between the 1959-60 group and the 1960-61 group did not attain the accepted level (95%), it did attain a level that is indicative of a trend (77%). This means that one can be 77% certain that the 2.48 standard score point growth in gain was due to the program in social studies rather than due to chance. Such a trend would support the efficacy of the program in social studies as against no program.

Though the significance in the growth of gains in reading achievement between the 1960-61 group and the 1961-62 group did not attain the accepted level (95%), it did attain a level (80%) that is indicative of a positive trend. See Table II. Ostensibly, one can assume that this indicates an 80% certainty that the 2.53 standard score point growth in gain was due to the developmental reading program being operative in both social studies and science rather than due to chance. This assumption is supported by the fact that the growth in reading achievement gains of the 1959-60 group compared to those of the 1961-62 group attained a level of significance beyond the accepted level (99%). This is supportive in that one can be 99% certain that the 5.01 standard score point growth in pre to post-test gain was due to the developmental reading program being operative in social studies and science rather than due to chance. Such a trend supported by such significance would underscore the efficacy of the program being operative in social studies and science as against only in social studies or with no program.

Caution must be exercised here, however. One cannot be <u>certain</u> that such significant gains were due to the fact that the program was operative in two subject areas rather than none or one. It is possible that the gains were due to the fact that the developmental reading program in social studies was in its second year of operation and the teachers involved had had an extra year's experience in adopting the methods and materials which are a part of the program. An examination of Tables VI and VII sheds some light on this point. One will note that growth in gains in science content achievement between the 1959-60 group and the 1961-62 group fell far short of the accepted level (40%). This would indicate that there is only a 40% certainty that the 1.87 raw score point growth was due to the program being operative in two subjects or, it indicates 60% certainty that this growth was

[illegible] only to chance. Contrasted to this is the growth of the gains in social studies content achievement between these two groups. Even though the level of significance (66%) did not reach the accepted level, it did attain a substantially higher level than that shown for science subject content. Though the 3.14 raw score point growth in gain is 34% certain to be due to chance, it is sufficiently less than the 60% chance factor of the science achievement. This difference indicates a trend regarding efficacy of teacher experience that should be investigated further.

This investigation of growth of achievement gain both in science content and in social studies content was accomplished by comparing a group that was exposed to the developmental reading program in both subject areas to a group that was exposed to no developmental reading program. One can say with certainty that the only variable involved here was teacher experience in the developmental reading program since the social studies teachers participated in the program during the 1960-61 school year while the science teachers did not. Therefore, it seems logical to assume that the highly significant growth in gains in reading achievement between these two groups and the substantially indicative growth in social studies content achievement as compared with limited growth in science content achievement, is due more to the fact that the social studies teachers had had an additional year's experience in the program than the fact that the program was operative in two subject areas.

This assumption substantiates the writer's observations of the operation of the program. It was obvious that the social studies teachers' additional experience gave them a greater facility with the methods and materials involved in the program. This would indicate the worth of continuing the program in science another year to note whether the added experience will allow a similar facility in the program and will, in turn, substantially effect the achievement level of the students both in reading skills and subject content.

By way of summary, one may note the following:

1. The instructional program in reading, science, and social studies is of sufficient quality to cause a significant level of achievement in skills development and acquisition of subject content regardless of the presence of the developmental reading program or, when it is present, regardless of the number of subjects in which it is operative.

2. Where the program is operative in two subject areas and the teachers in one of the subject areas have had a previous year's experience in the program, the gains in achievement in reading are significantly greater than where there is no developmental reading program. One cannot be certain as to whether the significance is due to the program being operative in a second area or due to the additional experience of some teachers in the program.

3. Indications are that the significant growth in gains in reading achievement is due more to the added experience of teachers in the program than it is due to the program's operation in a second subject area.

4. Though the accepted level of significance was not reached, indications are that growth in gains in subject content achievement is measureably effected by the developmental reading program being operative in the subject area.

5. Indications of improved achievement in subject content due to the program are sufficiently strong to support a recommendation that the program be continued in both social studies and science.

6. Indications of improved achievement in subject content and reading skills due to greater facility of teachers in the methods and materials related to the developmental reading program, support a recommendation that the program be continued in science to allow teachers to benefit from the additional experience.

7. Indications of success of the developmental reading program support a recommendation that a study be made of the advisability of extending the program into other subjects and into other grade levels.

Future Plans

Since this program is being adopted as a regular part of the district curriculum in seventh grade science, it is felt that it no longer qualifies for aid as an experiment. Therefore, no application for such aid will be made. However, evaluation will be continued in as many ways as possible. Teacher groups will meet in order to examine the program in the light of the above findings to determine areas of strength that can be capitalized upon and weakness that can be eliminated. The incoming seventh grade for the 1962-63 school year will be carefully studied and their performance compared with performances of previous years so as to take into account the effect of the science teachers' second year of experience in the program. Teachers of the two subject areas will meet at other grade levels to examine the materials, examine the program, and to make judgments as to the feasibility of extending the program into other subjects and grade levels.

TABLE I

Reading Achievement

-Comparing-

Pre to Post-test Standard Score
Point Gain by the Three Groups,
Based on the Iowa Silent Reading Test

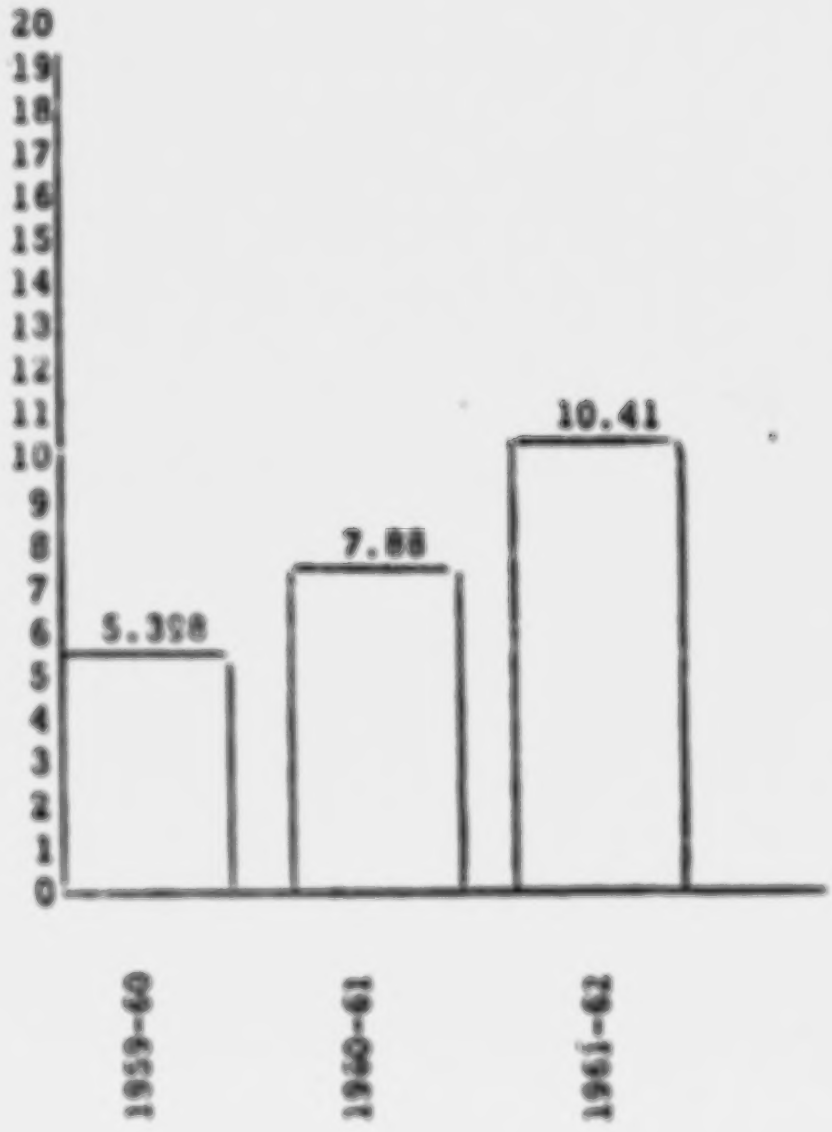

TABLE II

Reading Achievement

- Comparing -

Levels of Significance of Growth of
Pre to Post-test Gains, Between Various Groups,
Based on the Iowa Silent Reading Test

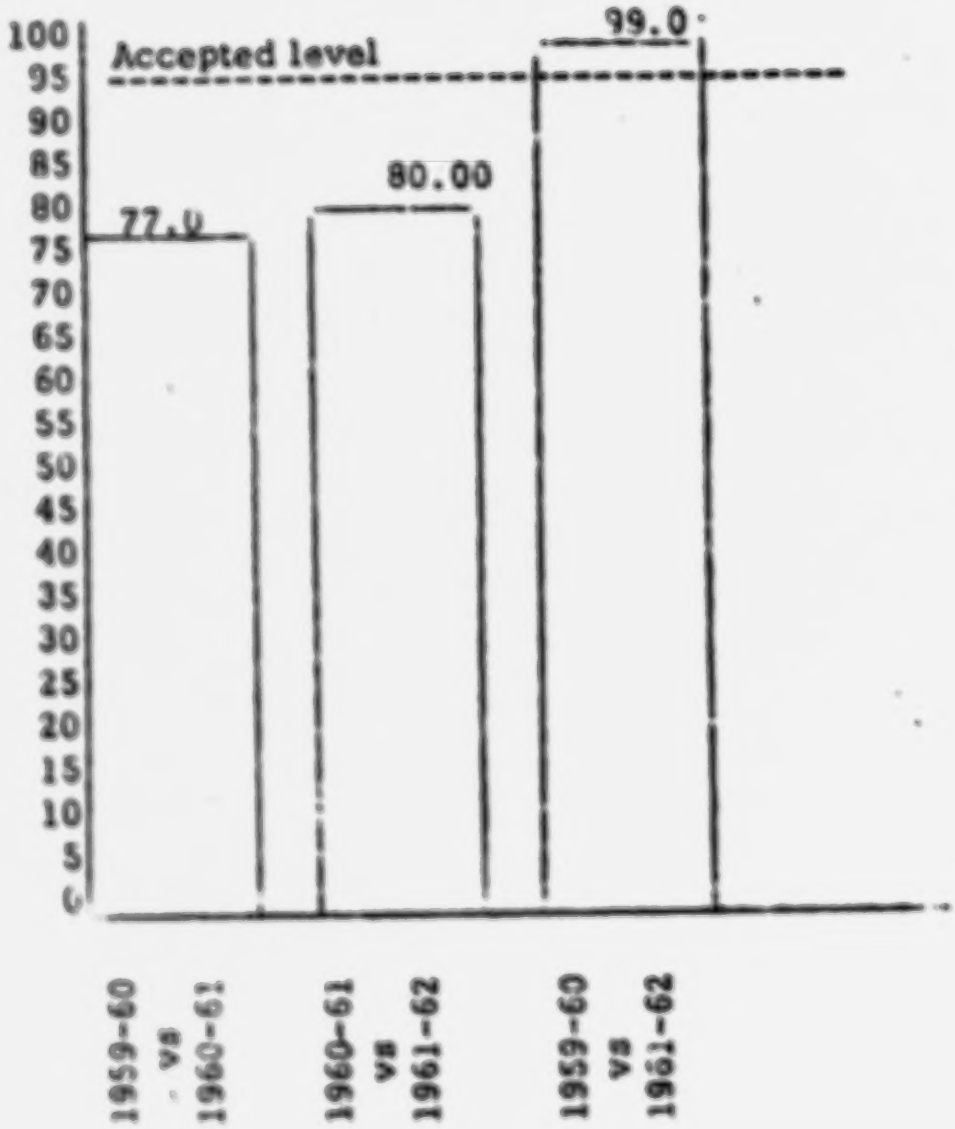

TABLE III

Reading Achievement

- Comparing -

Growth of Pre to Post-test Gain in
Standard Score Points, Between Various Groups,
Based on the Iowa Silent Reading Test

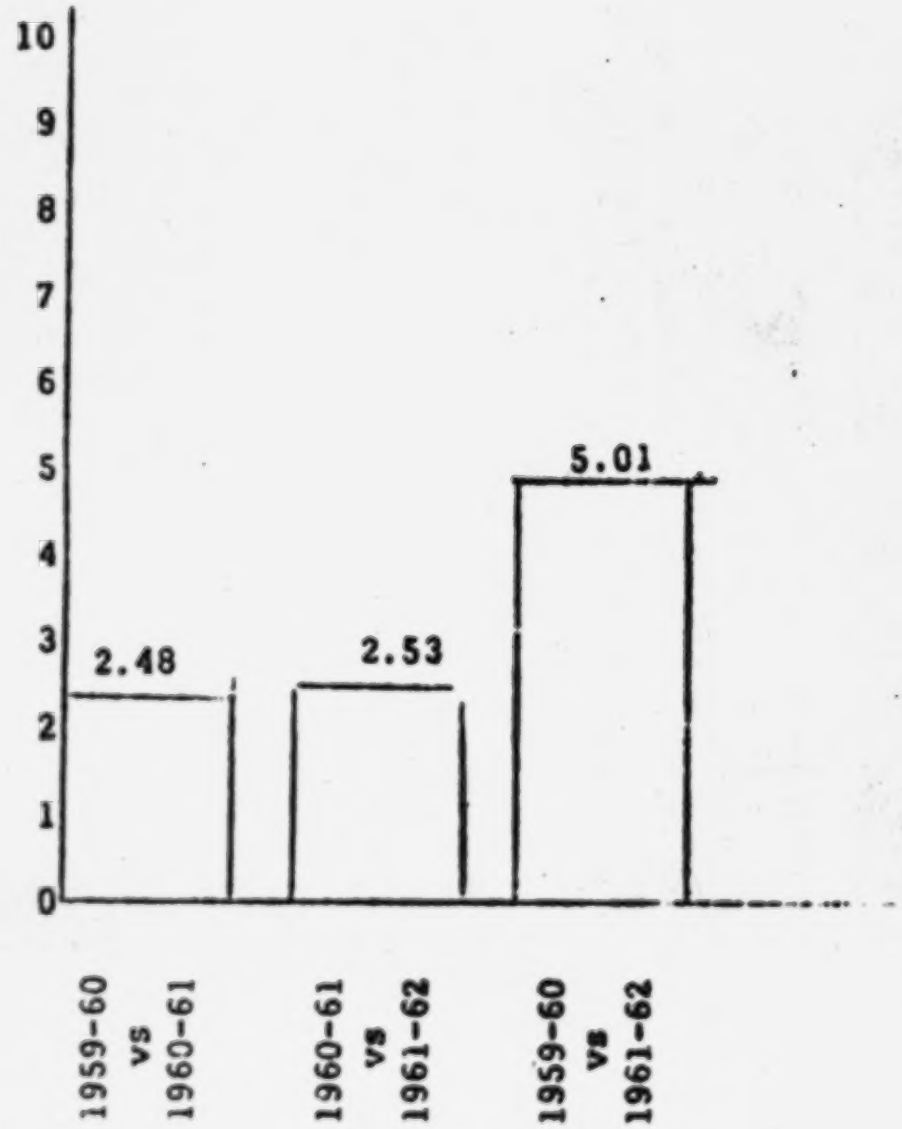

TABLE IV

Social Studies

- Comparing -

Pre to Post-test Raw Score Gain, By the Two Groups,
Based on the New York State Social Studies Survey Test

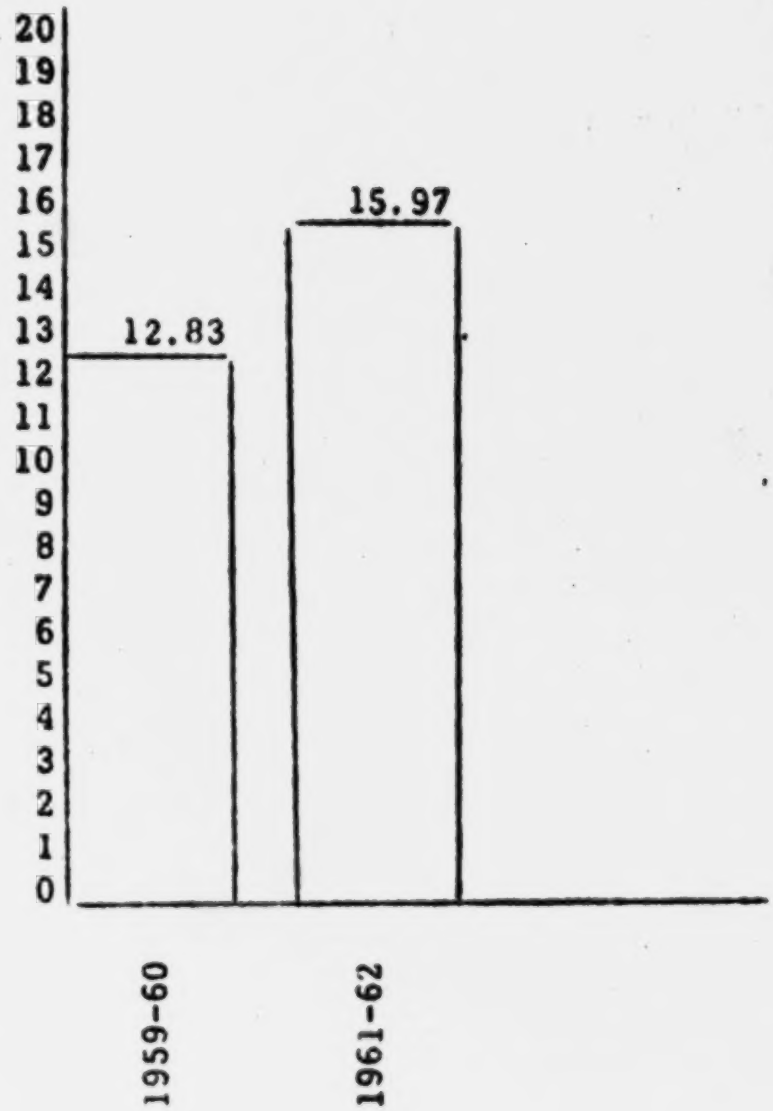

TABLE V

Science

- Comparing -

Pre to Post-test Raw Score Gain, by the Two Groups,
Based on the New York State Science Survey Test

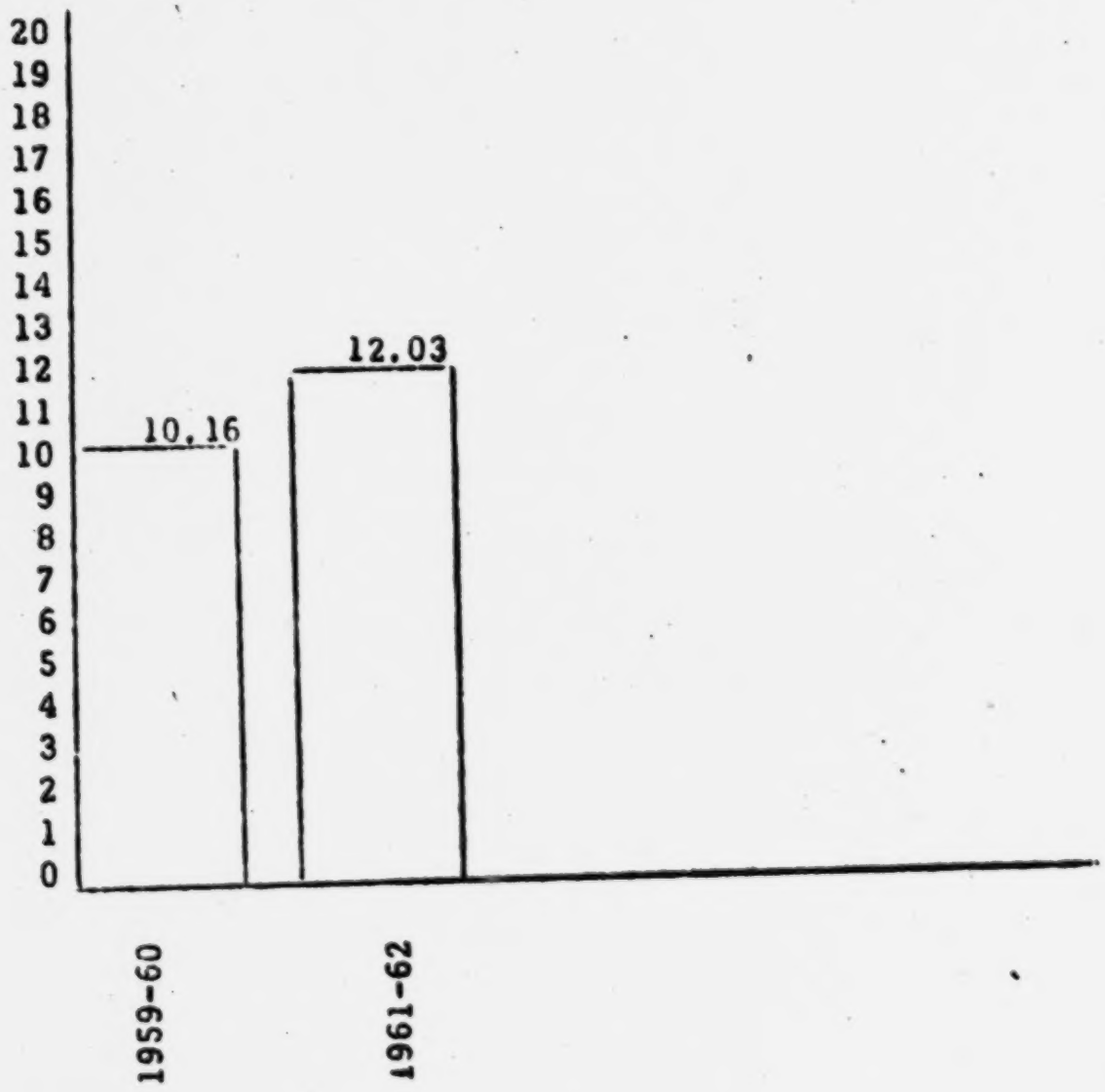

TABLE VI

Science and Social Studies

- Comparing -

Levels of Significance of Pre to Post-test Gains,
between 1959-60 and 1961-62 Groups
Both in Science and Social Studies
Based on the Respective Survey Tests

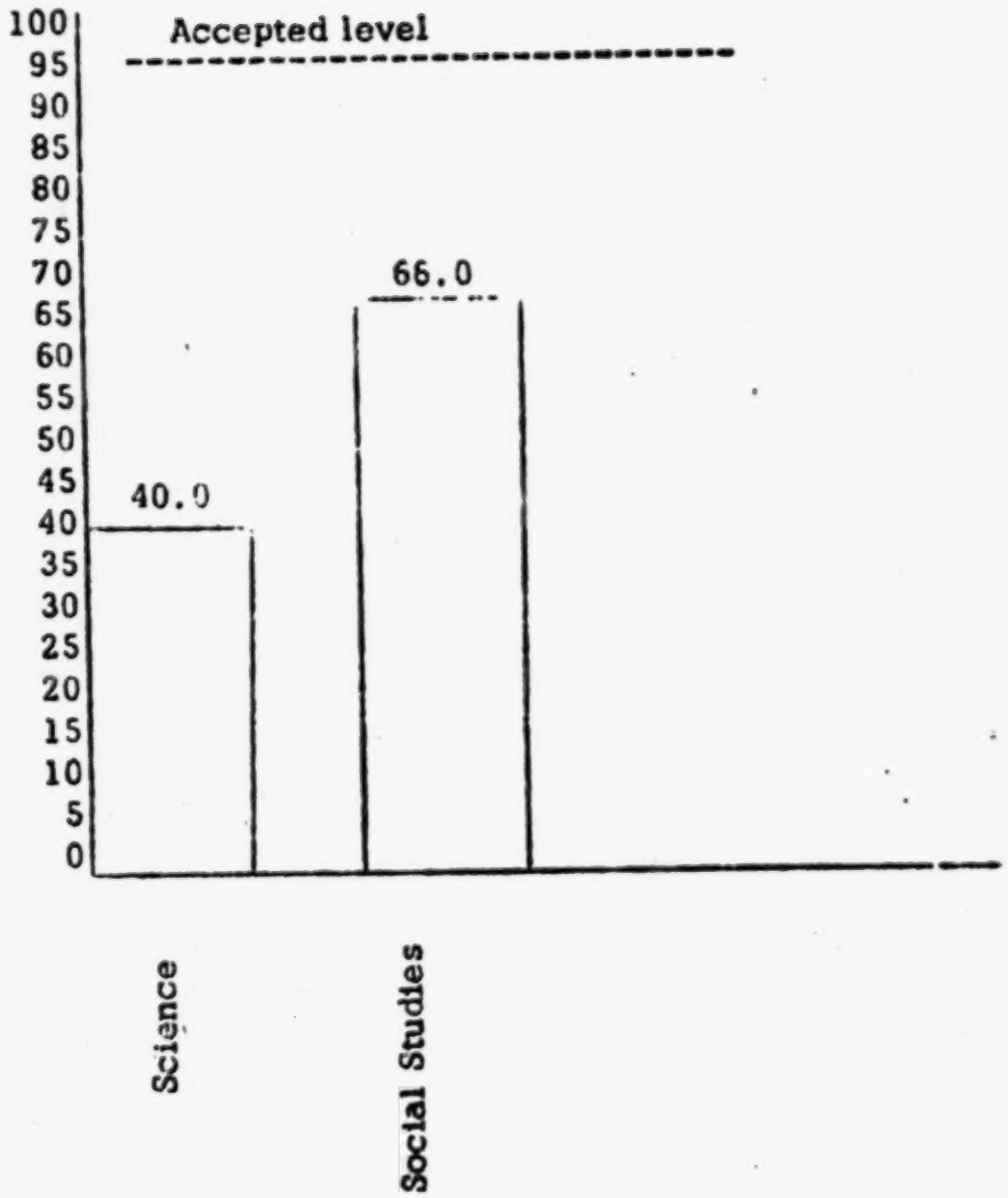

TABLE VII

Science and Social Studies

- Comparing -

Growth of Pre to Post-test, Gain in Raw Score
Between 1959-60 and 1961-62 Groups,
Both in Science and Social Studies,
Based on the Respective Survey Tests.

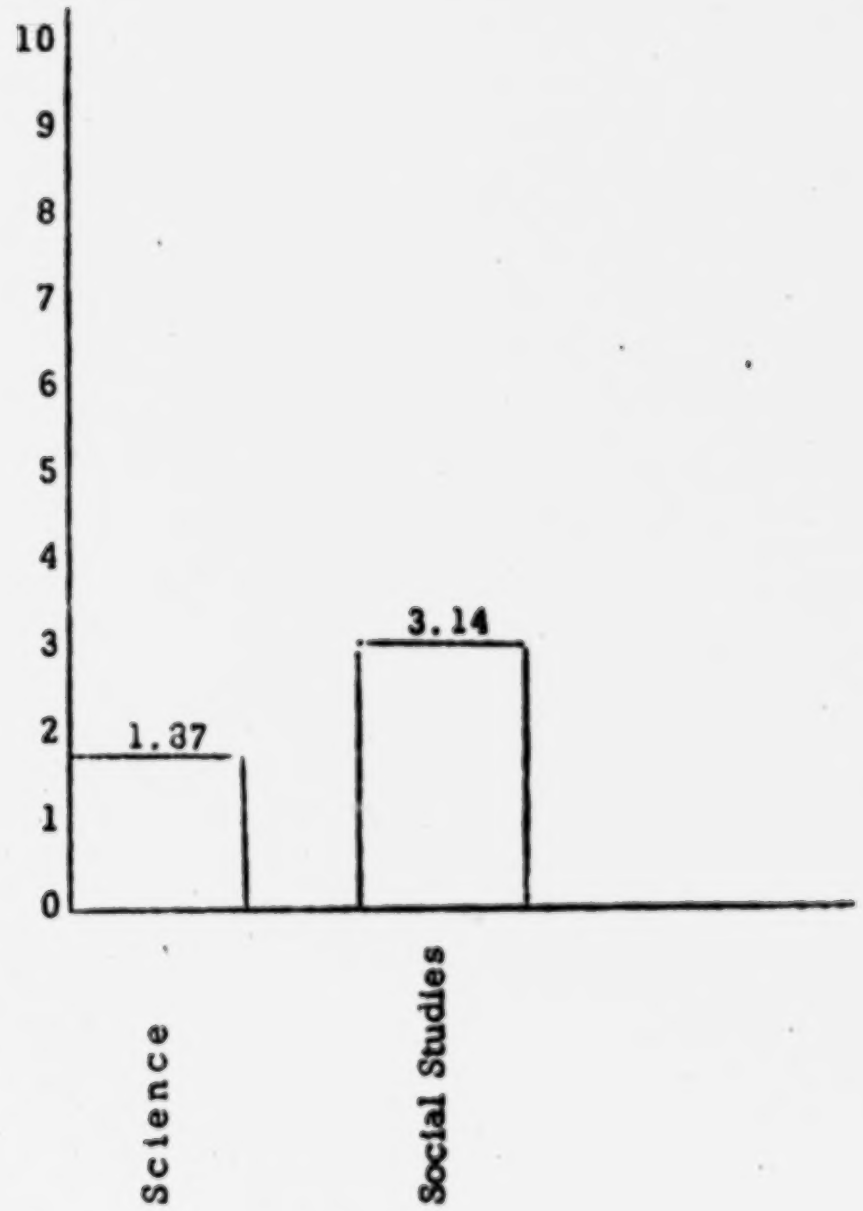

TABLE VIII

Years	Variate	Pupils	Σy	$\bar{y}=\frac{\Sigma y}{n}$	Σy^2	Σx	$\bar{x}=\frac{\Sigma x}{n}$	Σx^2	Σxy	Slope
1959-60	Pre ISR	98	15,819	161.42	2,570,133	-	-	-	-	-
	IC	98	-	-	-	10,627	108.44	1,174,703	-	-
	Regression	-	-	-	-	-	-	-	1,726,612	0.5025
1959-60	Post ISR	98	16,350	166.84	2,746,328	-	-	-	-	-
	IC	98	-	-	-	10,627	108.44	1,174,703	-	-
	Regression	-	-	-	-	-	-	-	1,785,455	0.5590
1960-61	Pre ISR	100	15,786	157.86	2,510,852	-	-	-	-	-
	IC	100	-	-	-	11,016	110.16	1,229,798	-	-
	Regression	-	-	-	-	-	-	-	1,750,858	0.7294
1960-61	Post ISR	100	16,574	165.74	2,764,138	-	-	-	-	-
	IQ	100	-	-	-	11,016	110.16	1,229,798	-	-
	Regression	-	-	-	-	-	-	-	1,836,680	0.6690
1961-62	Pre ISR	119	18,874	158.61	3,021,998	-	-	-	-	-
	IQ	119	-	-	-	13,411	112.70	1,545,143	-	-
	Regression	-	-	-	-	-	-	-	2,150,467	0.6936
1961-62	Post ISR	114	19,306	169.35	3,297,148	-	-	-	-	-
	IC	114	-	-	-	12,906	113.21	1,492,846	-	-
	Regression	-	-	-	-	-	-	-	2,209,381	0.7477
1959-60	Pre Soc.St	69	4,345	62.97	305,597	-	-	-	-	-
	IC	69	-	-	-	7,546	109.36	836,322	-	-
	Regression	-	-	-	-	-	-	-	488,282	1.1833
1959-60	Post Soc.St	69	5,250	76.09	436,754	-	-	-	-	-
	IQ	69	-	-	-	7,565	109.64	840,963	-	-
	Regression	-	-	-	-	-	-	-	589,974	1.2443

TABLE VIII (Cont'd)

Years	Variate	Pupils	Σy	$\bar{y}=\frac{\Sigma y}{n}$	Σy^2	Σx	$\bar{x}=\frac{\Sigma x}{n}$	Σx^2	Σxy	Slope
1961-62	Pre Soc.St	115	7,490	65.13	531,884	-	-	-	-	-
	IQ	115	-	-	-	12,990	112.96	1,505,572	-	-
	Regression	-	-	-	-	-	-	-	874,272	0.8486
1961-62	Post Soc.St	113	8,868	78.48	762,888	-	-	-	-	-
	IQ	113	-	-	-	12,749	110.43	1,471,531	-	-
	Regression	-	-	-	-	-	-	-	1,037,272	1,1076
1959-60	Pre Sci	83	5,910	71.20	463,258	-	-	-	-	-
	IQ	83	-	-	-	9,094	109.57	1,013,973	-	-
	Regression	-	-	-	-	-	-	-	665,856	1.0423
1959-60	Post Sci	83	6,758	81.42	594.614	-	-	-	-	-
	IQ	83	-	-	-	9,099	109.63	1,013,973	-	-
	Regression	-	-	-	-	-	-	-	759,387	1.1244
1961-62	Pre Sci	116	8,989	77.49	758,185	-	-	-	-	-
	IQ	116	-	-	-	13,088	112.83	1,508,644	-	-
	Regression	-	-	-	-	-	-	-	1,043,346	0.9118
1961-62	Post Sci	113	10,036	88.34	966,908	-	-	-	-	-
	IQ	113	-	-	-	12,729	112.65	1,465,679	-	-
	Regression	-	-	-	-	-	-	-	1,167,179	0.9402

APPENDIX

STATISTICAL REPORT

OF

TEACHING READING THROUGH SEVENTH GRADE SCIENCE CONTENT

DESIGN OF THE STATISTICAL ANALYSIS

The Iowa Silent Reading pretest for the three years: 1959-60, 1960-61, 1961-62; was examined for comparability using the chi^2 test for the mathematical model of the frequency distribution and the t test for equivalency. The New York State social studies and the New York State science were examined similarly. Furthermore, the general ability test (California Test of Mental Maturity) for the three years also was examined similarly. The results of calculations did not permit the acceptance of comparability, so classical variance analysis of the three post tests for the three years could not be used.

The necessary calculations of the linear regressions of the pre-tests on the IQ were performed. The type of the distribution of the residuals was found by the chi^2 test. Again, the results of the calculations of the parameters of the regressions did not permit the acceptance of comparability, so simple regression analysis of the post test on IQ could not be used.

The set of regressions of pre-test on IQ and post test on IQ for the three years was used in the analysis. Comparisons were made by the slopes and the "intercepts".

In the interpretation of the design, three words were used: slope, "gain", and "growth" which are defined as follows:

Slope is the amount of increase in test score relative to an increase in IQ (the usual interpretation of a slope).

"Gain" is the difference between a pre-test regression and a post-test regression, if the regressions are parallel.

"Growth" is the difference of two "gains" whose regressions are also parallel.

PROCEDURE

Form the least squares linear regressions with independent variable of IQ and dependent variable of either pre test or post test as listed below for the available data.

Years	ISR		Test Social Studies Survey		Science Survey		Independent Variable
	Pre	Post	Pre	Post	Pre	Post	
1959-60	x	x	x	x	x	x	IQ
1960-61	x	x	-	-	-	-	IQ
1961-62	x	x	x	x	x	x	IQ

The equation for the general regression (the expected value of the dependent variable, y) is as follows:

$$E(y) = \bar{y} + b(x - \bar{x})$$

where

$\bar{y}$ and $\bar{x}$ are the means of y and x

and

$$b = \frac{1}{n-1} \sum_{i=1}^{n} \frac{(y_i - \bar{y})(x_i - \bar{x})}{s_x^2}$$

where

s_x^2 is the variance of x

and

n is the number of pupils

For each pupil, the residual, d_i is

$$d_i = y_i - E(y_i)$$

where

$\bar{d}$ is the mean of d

and

s_d^2 is the variance of d

Form the theoretical normal frequency distribution of p (d). The consistency of the actual frequency distribution of d_i to the theoretical normal frequency distribution of d is tested by the chi^2 test through an hypothesis.

$$H_1: \sum_{i=1}^{k} \frac{(f_i - n\,p_i(d))^2}{n\,p_i(d)} \leq chi^2$$

where

k = number of intervals

f_i = actual frequency of d_i in one interval

n = total residuals

$p_i(d)$ = theoretical normal frequency in one interval

chi^2 = cumulative frequency distribution integrated so that $P(chi^2)$ is the probability in the interval (0,1)

In this study $P(chi^2) = 0.95$ or the 0.05 level

Accept H_1 $\leq$ chi^2

Reject H_1 $>$ chi^2

For $P(chi^2) = 0.95$

It is sufficient to consider only the residuals of the pre test regressions for the reliability of the samples. In each regression there were two independent randomly picked samples, with replacement

$$d_i = d_{i1} \neq d_{i2}$$

$$n = n_1 \neq n_2$$

Calculate the two means, $\bar{d}_1$ and $\bar{d}_2$, and the two variances, $s^2_{\bar{d}1}$ and $s^2_{\bar{d}2}$. The reliability can then be compared by an hypothesis.

$$H_2: \left| \frac{\bar{d}_1 - \bar{d}_2}{\sqrt{\frac{s^2_{\bar{d}_1}}{n_1} + \frac{s^2_{\bar{d}_2}}{n_2}}} \right| \leq 1$$

In the accumulation of the data it was found that some of the pupils' IQ scores were not CTMM scores but were the New York State Aptitude Test or the Otis IQ test. The specific IQ's were uncorrelated with the Junior High Schools. Approximate conversions were established. It is sufficient to test two residuals using variance analysis for the three IQ tests (CTMM and the two conversions.) This comparison can be tested for consistency by an hypothesis.

H_3 : $\frac{s_2^2}{s_1^2}$ $\leq$ F

where $F_{\frac{1}{2}}$ variance ratio

Accept $\leq$ F

Reject $>$ F

For P (F) = 0.95

where

$$s_2^2 = \frac{1}{k-1}\sum_{k=1}^{3} n_k (\bar{d} - \bar{\bar{d}})^2 \qquad ; f = 2$$

$$s_1^2 = \frac{1}{n-k-2}\left(\underbrace{\sum_{i=1}(d_i - \bar{d})}_{\text{within sets}} - \underbrace{\sum^{3} n_k(\bar{d}_k - \bar{\bar{d}})^2}_{\text{without sets}}\right) \qquad ; f = n-5$$

Consider the set of the regressions in one test for three years (ISR) and the set of the regressions in one test for two years. (Social Studies or Science Survey.) The analysis of the slopes and "intercepts" follows according to the method of Hold, Statistical Theory with Engineering Applications, p.579, John Wiley, 1955. First the variance of the regressions ($s^2_{y/x}$) must be tested for equality using the Bartlett Test. If the variances are not equal, then a variation of the analysis is used.

Consider the slopes of the regressions in one set. To compare the regressions for equivalency, state this hypothesis.

H_4: $\frac{s_2^2}{s_1^2}$ $\leq$ F

where F is the variance ratio

Accept $\leq$ F

Reject $>$ F

For P (F) = 0.95

where

$$s_2^2 = \frac{1}{m-1} \sum_{u=1}^{m} (b_u - \bar{b})^2 (n_u - 1) s_{x_u}^2 \quad ; f = m - 1$$

and

$$s_1^2 = \frac{\sum_{u=1}^{m} (n_u - 2) s_{y_u/x_u}^2}{\left(\sum_{u=1}^{m} n_u \right) - 2m} \quad ; f = \text{denominator of } s_1^2$$

and

$$\bar{b} = \frac{\sum_{u=1}^{m} b_u (n - 1) s_{x_u}^2}{\sum_{u=1}^{m} (n_u - 1) s_{x_u}^2}$$

where

u is the number of regressions in the set.

If H_4 is rejected, and if the Type II probability is also rejected, then the slopes are not equivalent and an immediate interpretation follows or further development of the analysis is required.

If H_4 is accepted, the analysis of the "intercept" follows for parallel slopes

To test the identity of two parallel regressions, state

$$H_5 : \frac{D}{\sqrt{s_d}} \quad \text{L} \quad t$$

where t is the t distribution

and

$$D = \bar{E}(y_i) - \bar{E}(y_j)$$

$$= \bar{y}_i - \bar{y}_j - b(\bar{x}_i - \bar{x}_j)$$

where
b is as in H_4

and

$$S_D^2 = s_o \left(\frac{1}{n_i} + \frac{1}{n_j} + \frac{(\bar{x}_i - \bar{x}_j)^2}{\sum_{u=1}^{m} (n_u - 1)\, s_{x_u}^2} \right)$$

$$s_o^2 = \frac{\sum_{u=1}^{m} (n_u - 2)\, s_{y_u/x_u}^2 + \sum^{m} (n_u - \bar{b})\,(n_u - 1)\, s_{x_u}^2}{\left(\sum_{u=1}^{m} n_u \right) - m - 1}$$

Accept H_5 $\leq$ t

Reject H_5 $>$ t

For P (t) = 0.95

For one pair of regressions, e.g. pre and post tests in one year, if H_4 is rejected, then this is a measure of the "gain". The difference of two "gains" can be measured by H_{4_1} (by the addition theorem). This difference is a measure of the "growth."

The combination of the tests of significance use the u test of Hold (p407) and the omnibus Fischer Test (p 407, ibid) for further analysis.

RESULTS

The residuals of the 14 regressions were tested for consistency to the theoretical normal distribution (p= 0.95), H_1. Ten fractiles were used and the results are tabulated below:

ISR	Actual chi^2	Theoretical chi^2	d.f.
1959-60			
pre	3.09	9.49	4
post	6.14	9.49	4
1960-61			
pre	2.09	7.81	3
post	3.50	11.1	5
1961-62			
pre	2.99	9.49	4
post	3.69	11.1	5

	Actual chi	Theoretical chi^2	d.f
Social Studies Survey			
1959-60			
pre	3.69	9.49	4
post	7.51	9.49	4
1961-62			
pre	6.52	9.49	4
post	3.07	11.1	5
Science Survey			
1959-60			
pre	13.26	9.49	4
post	6.75	9.49	4
1961-62			
pre	6.15	11.1	5
post	5.06	11.1	5

Only one of the residuals frequency distributions was rejected for consistency to the theoretical normal frequency distribution, H_1. Consistency was not rejected by the power test (Type II) since it was 50-60%. Moreover, this pre test residual distribution was rectified in the post test residual distribution and, therefore, it can be assumed that the residuals of all 14 regressions were distributed normally.

The reliability of the samples chosen was tested by H_2 for the pre test residuals. The results are tabulated below:

	Actual t	Theoretical t	d.f.
ISR			
1959-60	1.31	1.98	94
1960-61	0.76	1.98	96
1961-62	0.24	1.98	110
Social Studies Survey			
1959-60	1.40	1.99	65
1961-62	0.88	1.98	103
Science Survey			
1959-60	1.29	1.99	79
1961-62	0.34	1.98	112

The comparison of the two IQ conversions to CTMM-IQ was checked by the H_3 hypothesis for the pre ISR 1959-60 residuals and was accepted since the actual F of 1.93 was less than the theoretical F of 3.10. The comparison of the two IQ conversions to CTMM-IQ was checked by the H_3 hypothesis for the pre Social Studies Survey 1959-60 residuals and was accepted since the actual F of 0.752 was less than the theoretical F of 3.14.

Consider the slopes of the six regressions of ISR. The Bartlett Test showed that the six variances were equivalent since the actual chi^2 of 2.95 was less than the theoretical chi^2 of 11.1. The slopes were compared for equivalency using the H_4 hypothesis. The actual F of 2.17 was less than the theoretical F of

2.22 for $f_1 = 5$, $f_2 = 617$ and, therefore, all of the slopes are parallel.

The "gain" in the pre to post test regressions of ISR 1959-60 was 5.398 in ISR units. Using the H_5 hypothesis, this "gain" was shown to be significant since the actual t of 3.67 was greater than the theoretical t of 1.96. The two tailed power test (Type II) was P = 0.95. The "gain" was significant and decisive assurance had been obtained.

The "gain" in the pre to post test regressions of ISR 1960-61 was 7.880 in ISR units. Using the H_5 hypothesis, this "gain" was shown to be significant since the actual t of 5.408 was greater than the theoretical t of 1.96. The two tailed power test (Type II) was greater than P=0.999. The "gain" was significant and decisive assurance had been obtained.

The "gain" in the pre to post test regressions of ISR 1961-62 was 10.41 in ISR units. Using the H_5 hypothesis, this "gain" was shown to be significant since the actual t of 7.71 was greater than the theoretical t of 1.96. The two tailed power test (Type II) was greater than P= 0.999. The "gain" was significant and decisive assurance had been obtained.

The "growth" of the "gains" in 1960-61 and the "gains" in 1959-60 was 2.462 in ISR Units. This "growth" was not significant using the H_5 hypothesis since the actual t of 1.189 was less than the theoretical t of 1.97.

The "growth" of the "gains" in 1961-62 and the "gains" in 1959-60 was 4.99 in ISR units. This "growth" was significant using the H_5 hypothesis since the actual t of 2.50 was greater than the theoretical t of 1.97. The power test (Type II) was P = 0.70. This "growth" was significant but not with decisive assurance.

Furthermore, the two "growth" results of the ISR, since they are stochastically independent, can be combined by the u test of Hold and Fischer's onmibus test of combined significance. The u test showed a P=0.993 and the Fischer omnibus test, which is a chi^2 test, showed a P=0.95 since the actual chi^2 of 11.96 was greater than the theoretical chi^2 of 11.1

Consider the slopes of the four regressions of the Social Studies Survey. The Bartlett Test showed that the four variances were equivalent since the actual chi^2 of 5.53 was less than the theoretical chi^2 of 7.81. The slopes were compared for equivalency using the H_4 hypothesis. The actual F of 3.016 was greater than the theoretical F of 2.63 for $f_1 = 3$, $f_2 = 358$. However, the power test (Type II) is not significant at the 95% level since the theoretical F number was 3.16. The slopes were considered parallel, at least for a first approximation.

The "gain" in the pre to post test regressions of the Social Studies Survey 1959-60 was 12.83 in the Social Studies Survey units. Using the H_5 hypothesis, this "gain" was shown to be significant since the actual t of 4.93 was greater than the theoretical t of 1.96. The two tailed power test (Type II) was P=0.997.

The "gain" was significant and decisive assurance had been obtained.

The "gain" in the pre to post test regressions of the Social Studies Survey 1961-62 was 15.97 in the Social Studies Survey units. Using the H_5 hypothesis, this "gain" was shown to be significant since the actual t of 7.87 was greater than the theoretical t of 1.96. The two tailed power test (Type II) was greater than P=0.999. The "gain" was significant and decisive assurance had been obtained.

The "growth" of the "gains" in 1961-62 and the "gains" in 1959-60 was 3.136 in Social Studies Survey units. This "growth" was not significant using the H_5 hypothesis since the actual t of 0.947 was less than the theoretical t of 1.97. This result is indicative only because the sign is positive.

Consider the slopes of the four regressions of the Science Survey. The Bartlett Test showed that the four variances were equivalent since the actual chi^2 of 1.60 was less than the theoretical chi^2 of 7.81. The slopes were compared for equivalency using the H_4 hypothesis. The actual F of 0.679 was less than the theoretical F of 2.62 for $f_1 = 3$, $f_2 = 387$ and, therefore, all of the slopes are parallel.

The "gain" in the pre to post test regressions of the Science Survey 1959-60 was 10.16 in Science Survey units. Using the H_5 hypothesis, this "gain" was shown to be significant since the actual t of 3.78 was greater than the theoretical t of 1.97. The two tailed power tost (Type II) was P = 0.967. The "gain" was significant and decisive assurance had been obtained.

The "gain" in the pre to post test regressions of the Science Survey 1961-62 was 12.03 in Science Survey units. Using the H_5 hypothesis, this "gain" was shown to be significant since the actual t of 5.26 was greater than the theoretical t of 1.97. The two tailed power test (Type II) was greater than 0.999. The "gain" was significant and decisive assurance had been obtained.

The "growth" of the "gains" in 1961-62 and the gains in 1959-60 was 1.875 in Science Survey units. This "growth" was not significant using the H_5 hypothesis since the actual t of 0.532 was less than the theoretical t of 1.97. The result is indicative only because the sign is positive.

CONCLUSIONS

The linear regressions of pre and post tests of Iowa Silent Reading on IQ had the same slopes for the three years for which there was data.

The linear regressions of pre and post tests of the New York Science Survey on IQ had the same slopes for the two years for which there was data.

The linear regressions of pre and post tests of the New York Social Studies Survey on IQ had the same slopes for the two years for which there was data, at least for a first approximation.

All of the "gains" of post test compared to pre test for the seven tests (ISR for 3 years, NY Science Survey for 2 years, NY Social Studies Survey for 2 years) were significant.

The "growth" in Iowa Silent Reading of the seventh grade 1961-62 compared to the seventh grade 1959-60 was significant and amounted to 5 points in ISR units.

The "growth" in Iowa Silent Reading of the seventh grade 1960-61 compared to the seventh grade 1959-60, and the "growth" in Iowa Silent Reading of the seventh grade 1961-62 compared to the seventh grade 1960-61 were not sign ificant, but were indicative.

The combined tests of significance showed the experimental program significant in the Iowa Silent Reading.

The "growth" of seventh grade New York Social Studies Survey 1961-62 compared with 1959-60 was 3.13 in New York Social Studies units and the "growth" of seventh grade New York Science Survey 1961-62 compared with 1959-60 was 1.87 in New York Science Survey units. These "growths" were not significant but were indicative only because of sign.